May 18-19, 2020

Politecnico di Milano

POLITECNICO DI MILANO





Applied Statistics 2019-2020
Permutation Tests

Simone Vantini

MOX, Dept. of Mathematics, Politecnico di Milano
Leonardo Campus, Building 14, Floor VI,
02 2399 4584
simone.vantini@polimi.it

Gaussian Assumption and Parametric Tests for the Mean(s)

p = # random variables (features)

$1 = p < n = \infty$

Thanks to the Central Limit Theorem, Gaussianity is not a key point.

$1 = p < n \leq \infty$

The *t*-distribution is meant to model situations in which the sample size is not very large. So the Gaussianity of data is required for the *t*-test. Univariate Gaussianity is, anyhow, not difficult to assess (normality tests).

$1 \leq p < n \leq \infty$

Hotelling's T^2 test rely on multivariate Gaussianity of data. If p increases, multivariate Gaussianity can be difficult to assess (curse of dimensionality).

$1 \leq n < p \leq \infty$

High-dimensional tests rely on the multivariate Gaussianity of data, and they are not robust with respect to the violation of Gaussianity.
Powerful Gaussianity tests are not available in the high-dimensional setting.

$1 \leq n < p = \infty$

In the functional case, normality is basically an unverifiable assumption.

Gaussian Assumption and Parametric Tests for the Mean(s)

All parametric tests (for the means) are exact
either **asymptotically** or
under the **Gaussianity assumption**

and

not exact otherwise

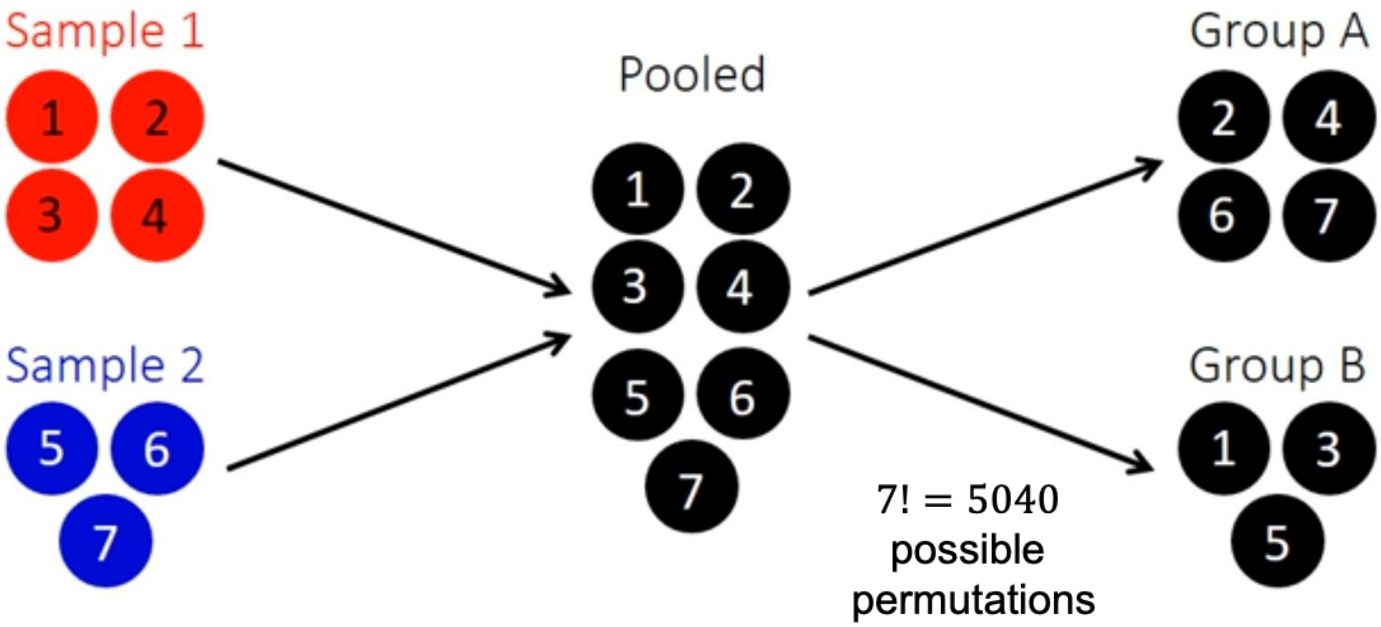
At the beginning of permutation tests

Let us suppose, for example, that we have measurements of the stature of ***a hundred Englishmen and a hundred Frenchmen****. It may be that* ***the first group are, on the average, an inch taller than the second****, although the two sets of heights will overlap widely. [...] The simplest way of understanding quite rigorously, yet without mathematics, what the calculations of the test of significance amount to, is to consider what would happen* ***if our two hundred actual measurements were written on cards, shuffled without regard to nationality, and divided at random into two new groups of a hundred each****. This division could be done in an enormous number of ways, but though the number is enormous it is a finite and a calculable number. We may suppose that* ***for each of these ways the difference between the two average statures is calculated****. Sometimes it will be less than an inch, sometimes greater.* ***If it is very seldom greater than an inch, in only one hundredth, for example****, of the ways in which the sub-division can possibly be made, the statistician will have been right in saying that* ***the samples differed significantly****. For* ***if,*** *in fact****, the two populations were homogeneous, there would be nothing to distinguish the particular subdivision*** *in which the Frenchmen are separated from the Englishmen* ***from among the aggregate of the other possible separations*** *which might have been made. Actually, the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method.*

Fisher, R. A. (1936). The coefficient of racial likeness and the future of craniometry, *Journal of the Anthropological Institute of Great Britain and Ireland*, pp. 57-63.

The two-population T-test in a glance

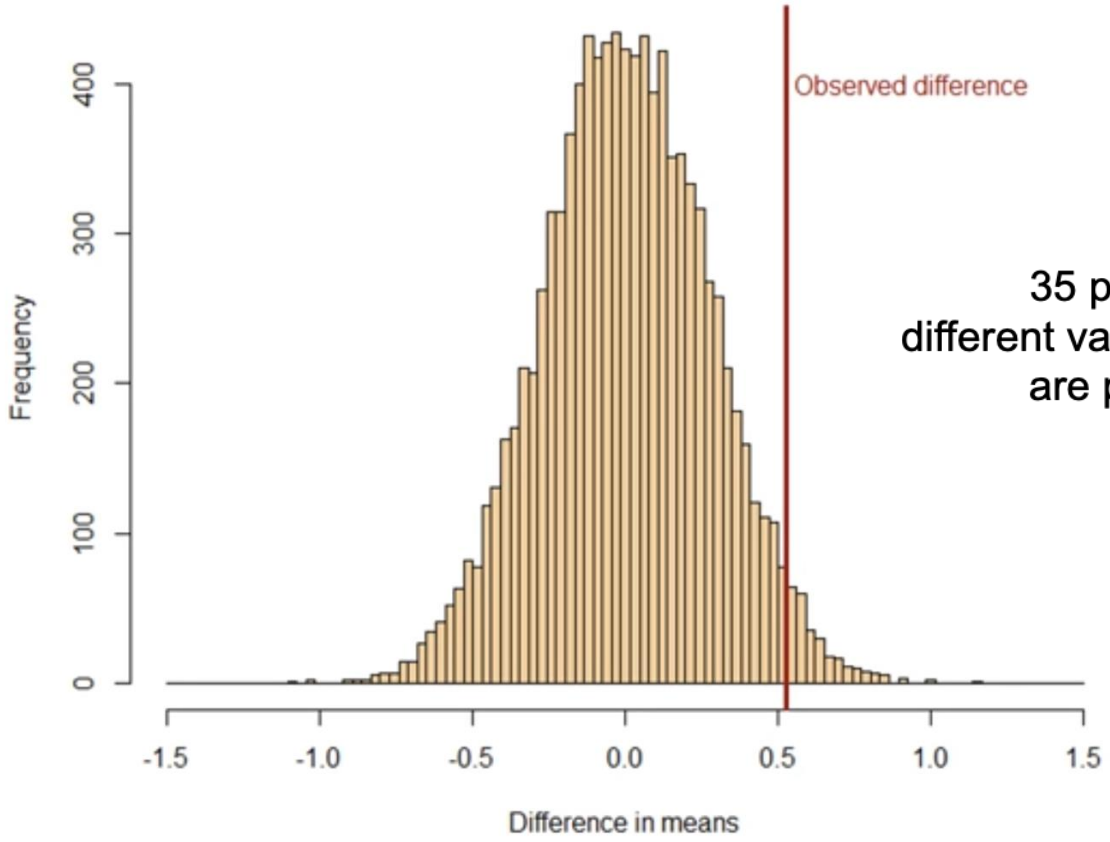
H0: $m_1 = m_2$ vs H1: $m_1 > m_2$



$\binom{7}{3} = 35$ possible different groupings

The two-population T-test in a glance

Permutational distribution of $T = \bar{x}_1 - \bar{x}_2$ **under the H0**



35 possible different values of $\bar{x}_1 - \bar{x}_2$ are possible

Conditional distribution of $T = \bar{x}_1 - \bar{x}_2$ given the pooled sample:

- Under the H0 all the 35 values of $\bar{x}_1 - \bar{x}_2$ are equally probable
- Under the H1 larger values of $\bar{x}_1 - \bar{x}_2$ are more probable

$$p = \frac{\sum_{k=1}^{35} I(T_k^* \geq T_0)}{35}$$

Permutation Tests

Some **corner-stones of permutational inference**:

- **Their aim is making fewer assumptions as possible on data distribution**

- **How they work**:
  - Likelihood-invariant transformations under the H0
    (Conditional inference within induced equivalence classes)
  - Selection of the test statistic:
    - no a-priori optimal test statisticits distribution has to be stocastically larger under the «targeted» H1 than under the H0
    - possibility of working in purely metric spaces (i.e., complex data)

- **Inferential properties**:
  - Finite-sample exactness (differently from bootstrap)
  - Consistency (if the test statistic is properly chosen)
  - Asymptotic equivalence to parametric tests (when the same test statistic is used and the parametric assumptions hold)

- **Large computational cost**s (Conditional Montecarlo)

Likelihood-invariant transformations

Two-population test and 1-way ANOVA:

→ Value permutations (equivalent to group labels permutation)

One-population test and and paired two-popolution test:

→ Recentering in H0 and sign swaps (assuming simmetry)

Independence test:

→ Pair Recoupling

“F-test” for linear models (linear regression and multi-way ANOVA)

→ Response permutations

“T-test” for linear models (linear regression and multi-way ANOVA)

→ Permutations of residuals of restricted model [*asymptotic*]

→ Permutations of residuals of complete model [*asymptotic*]